Appendix

Appendix A: model predictors

Word exposure index

Lexical frequencies were extracted from the English corpora of the CHILDES database (MacWhinney, 2000; Sanchez et al., 2019). Using the corresponding lexical frequencies from the Catalan and Spanish corpora was not possible due to the low number of Catalan participants and tokens available in those languages. Available words in the English corpora were mapped to their Catalan and Spanish translations (see Fourtassi et al., 2020 for a similar approach), and transformed to Zipf scores (Van Heuven et al., 2014; Zipf, 1949). Responses to words with missing lexical frequencies were excluded from analyses of the total number of items. Participant degree of exposure to the words' language were calculated as the percentage of exposure to the language, as reported by caretakers in the language profile section of the questionnaire. For example, for a participant with 90% exposure to Catalan, and 10% to Spanish, the DoE of the Catalan word *taula* would be 90%, and the DoE for the Spanish word *mesa* would be 10%.

We created a predictor (Exposure) to account for the exposure rate of each child to each of the words their parents responded to, weighted by the child's exposure to the language the word belongs to. the exposure of the i-th child to the j-th word measure is the product of the lexical frequency of the word (Zipf score) and the child's degree of exposure to the corresponding language (a proportion) (see Equation 1).

$$\text{Exposure}_{ij} = \text{Frequency}_i \times \text{DoE}_j \tag{1}$$

For instance, for a child who is reportedly exposure to Catalan 80% of the time, and to

Spanish 20% of the time, the expected exposure to the word *cavall* (horse, in Catalan, with a lexical frequency of) would be , while that of its translation to Spanish *caballo* would be .

Levenshtein similarity

Generally, a pair of words from two different languages that share meaning—translation equivalents—are considered as *cognates* if they share etymological origin (e.g., *table* and *taula*, in English and Catalan). The fact that two words share etymological origin frequently leads to them also being form-similar, as reflected by their overlapping orthographic or phonological form. This makes many cognates perceptually similar. In psycholinguistics, the term *cognateness* is often used to refer to the form-similarity that a pair of translation equivalents share, which has been found to impact how bilinguals process such word-forms (e.g., Costa et al., 2000; Spivey & Marian, 1999). Whether two form-similar translation equivalents are etymologically related or not is arguably tangential to the question of how cognateness affect language processing. For instance, the translation pair *much* and *mucho*, in English and Spanish have the same meaning, and are considerably similar at the orthographical and phonological level. Nonetheless, they come from different Proto-Indo-European roots (Campbell & Mixco, 2007). This does not keep bilinguals from processing such kind of word-forms word forms differently than other translation equivalents with no form-similarity like 'dog' and *perro*, in English and Spanish. In the scope of this experiment, we consider any form-similar translation equivalent as a *cognate* pair, regardless of whether both word-forms share etymological origin.

Our study capitalises in phonology, therefore we operationalised *cognateness* as the phonological similarity between a pair of translation equivalents. Our dataset contained 302 translation equivalents, each consisting in two word-forms: Catalan and Spanish (see Table 1 for an example). Measuring the phonological similarity between two word forms is non trial. While orthographic similarity can be calculated using string similarity/distance measures on the written word-forms, phonology poses additional challenges. A first, bottom-up approach to to the task might be considering acoustic similarity as a proxy to phonological similarity. This may involve registering audio recordings of a talker reading each word-form aloud, and then comparing their spectrograms, cochleograms, or formant tracks (e.g., Heeringa

& Gooskens, 2003). While this is possible, this method computes acoustic similarity, as opposed to phonological similarity. This is not optimal, since a purely acoustic measure of similarity ignores how the actual signal is perceived by a actual listener. Since approximately 12 months of age, humans perceive the acoustic sounds in speech as phonemes, according to the phonology of their native language(s) (Kuhl et al., 1992; Werker & Tees, 1984). This means that two listeners may not perceive the same linguistically relevant acoustic signal identically. This is why phonological similarity needs to be computed from discrete units that are meaningful for language perception.

A way to achieve this is to use phonological transcriptions of the word forms. Phonological transcriptions are symbolic, visual representations of the sounds in speech that capture some minimal set of its features following some standard. The International Phonetic Alphabet (IPA) is one instance of such standards, providing a set of symbols that correspond to specific phonemes, as defined by their position across a series of dimensions. Using IPA transcriptions of word forms, one can use the same similarity/distance metric as in the case of orthographic word-forms. This method has been successfully used to measure the pairwise similarity between phonological word-forms from the same language (e.g., Fourtassi et al., 2020) and across languages (Floccia et al., 2018; Heeringa, 2004), and also between orthographic word-forms (Schepens et al., 2012).

Multiple algorithms have been created to compute the distance between two strings of characters. The Levenshtein distance being one of them. The this algorithm calculates the minimum number of edit operations (additions, deletions, and substitutions) that one string must go through to become identical to the other string (Levenshtein, 1966) (see Equation 2). For instance, /ˈka.za/ (*house*, in Catalan), would need to go through one substitution to become identical to its translation equivalent in Spanish /ˈka.sa/ (/z/ is replaced by /s/). Therefore, the Levenshtein distance between /ˈka.za/ and /ˈka.sa/ is one. For an easier handling of non-ASCII characters (fairly prevalent in Catalan and Spanish, e.g., á, ü, ñ), we used phonological transcriptions in X-SAMPA format, as opposed to IPA, for its computer-friendly set of symbols (Wells, 1995).

$$\operatorname{lev}(x,y) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0 \\ \min \begin{cases} \operatorname{lev}_{x,y}(i-1,j)+1 \\ \operatorname{lev}_{x,y}(i,j-1)+1 \\ \operatorname{lev}_{x,y}(i-1,j-1)+1_{a_i \neq b_j} \end{cases} & \text{otherwise} \end{cases} \tag{2}$$

Here, a and b are the character strings corresponding to the phonological transcriptions of two word-forms belonging to the same translation equivalent, where each character corresponds to one phoneme expressed as a symbol from the SAMPA alphabet. i and j are the length (i.e., number of phonemes) of the s and t strings respectively.

To measure the phonological similarity between the translation equivalents in our study, we computed a normalised versions of the Levenshtein distance between each pair of word-forms that accounts for the length of the strings. The rationale behind this correction is that longer word-forms are more likely to differ from their counterpart, compared to shorter word forms, and that this tendency does not correspond necessarily to a larger distance in terms of perception. This normalisation is achieved by dividing the Levenshtein distance by the length of the longest string, which leads to a distance metric that ranges from 0 to 1. This can be interpreted as a proportion of characters in the longest string that need to be edited in order for that string to become identical to the shorter string. A 0% normalised Levenshtein distance indicates that both word-forms are identical. A 50% normalised Levenshtein distance indicates that half of the phonemes in the longest word-form must be edited for both word-forms to become identical. A 100% normalised Levenshtein distance indicates that the two word-forms are completely different, as *all* phonemes in the longest word-form must be changed for it to become identical to the other.

Finally, since we are interested in *cognateness*, in order to easy the interpretation of the analyses we subtracted the normalised Levenshtein distance from one, so that the metric indicated the *similarity* between the each pair of word-forms, instead of their distance. Equation 3 shows the formula of this metric.

$$1 - \frac{\operatorname{lev}(a,b)}{\max\{i,j\}} \tag{3}$$

The R package `stringdist` (Loo, 2014) offers the `stringsim()` function to compute this normalised Levenshtein similarity measure, see Table 1 for more examples:

Table 1: Normalised Levenshtein similarity computed for three exemplars of translation pairs in our study

Catalan	Spanish	Levenshtein
porta /pOrt5/	puerta /pwe4ta/	0.50
taula /tawl5/	mesa /mesa/	0.00
cotxe /kOtS5/	coche /kotSe/	0.40

Appendix B: model details

We used multilevel ordinal regression to model the cumulative probability of a *No* response, a *Understands* response, or a *Understands and Says* response (Bürkner & Vuorre, 2019) using the *logit* link function. The likelihood of the cumulative probability distribution of the responses is defined by Equation 4.

$$\textbf{Likelihood:} \qquad y_{ij} \sim \text{Cumulative}(p_k) \tag{4}$$

where: y is an observed response ($y \in \{\text{No, Understands, Understands and Says}\}$), i is the participant index, j is the translation equivalent (TE) index, p_k is a probability ($p \in (0, 1)$) that indicates the threshold k ($k \in (1, 2)$) between two response categories in the latent distribution p_k is then estimated using a logit regression model as indicated in Equation 5.

To test our hypotheses, we included several predictors in the regression model as fixed effects: the main effects of participant age (Age), word-form length (Length), word exposure index (Exposure), and of phonological similarity (Cognateness), the two-way interactions Age × Exposure, Age × Cognateness, and Exposure × Cognateness, and the three-way interaction Age × Exposure × Cognateness. We also included crossed random effects for participants and translation equivalents to account for the repeated measures in our dataset—each participant provided responses to multiple translation equivalents, and each translation equivalent was responded to by multiple participants (Gelman et al., 2020). For both grouping variables, we included random intercepts, random slopes, and correlation parameters for all predictors were repeated measures were observed in our dataset (Barr et al., 2013), see (Equation 5).

Linear model:

$$
\begin{aligned}
\ln \frac{p_k}{1-p_k} = {} & (\beta_{0_k} + u_{0_{i_k}} + w_{0_{j_k}}) + \\
& (\beta_1 + u_{1_i} + w_{1_j}) \cdot \text{Age}_i + \\
& (\beta_2 + u_{2_i} + w_{2_j}) \cdot \text{Length}_{ij} + \\
& (\beta_3 + u_{3_i} + w_{3_j}) \cdot \text{Exposure}_{ij} + \\
& (\beta_4 + u_{4_i}) \cdot \text{Cognateness}_{ij} + \\
& (\beta_5 + u_{5_i} + w_{3_j}) \cdot (\text{Age}_i \times \text{Exposure}_{ij}) + \\
& (\beta_6 + u_{6_i}) \cdot (\text{Age}_i \times \text{Cognateness}_{ij}) + \\
& (\beta_7 + u_{7_i}) \cdot (\text{Exposure}_{ij} \times \text{Cognateness}_{ij}) \\
& (\beta_8 + u_{8_i}) \cdot (\text{Age}_i \times \text{Exposure}_{ij} \times \text{Cognateness}_{ij})
\end{aligned}
\tag{5}
$$

where i and j index the participant and translation equivalent (TE), β_{0_k} is the fixed coefficient of the regression model for the intercept of threshold k, u_{0_i} and w_{0_j} are the by-participant and by-TE adjustments for β_{0_k} (i.e., random intercepts), respectively, β_{1-8} are the fixed coefficients of the regression model for the predictors of interest, u_{1-8_i} and w_{1-3_j} are the by-participant and by-TE adjustments forβ_{1-8} (i.e., random slopes), respectively.

We used the Bayesian framework to estimate the parameters in our model. This involves using the Bayes theorem to compute a distribution (*posterior distribution*) that describes what values of each parameter in the model are more likely given the data (*likelihood*, see Equation 4), and previous knowledge about such distribution (*prior distribution*, see Equation 6) (McElreath, 2020). This posterior distribution not only informs about the most likely values of our regression coefficients of interest, but also about the uncertainty around such estimations. We used a weakly informative prior for our parameters, with the exception of the main effect of Age, for which we specified a strongly informative prior based on previous literature about how age affects the acquisition of words (see Equation 6).

Prior:

$$
\begin{aligned}
\beta_{0_k} &\sim \mathcal{N}(-0.25, 0.1) && \text{[Intercept/response category threshold]}\\
\beta_1 &\sim \mathcal{N}(1, 0.1) && \text{[Age population-level coefficient]}\\
\beta_{2-8} &\sim \mathcal{N}(0, 1) && \text{[Rest of population-level coefficients]}\\
u_{0-8_i} &\sim \mathcal{N}(0, \sigma_{u_{0-8_i}}) && \text{[Participant-level coefficient variability]}\\
w_{0-3_j} &\sim \mathcal{N}(0, \sigma_{w_{0-3_j}}) && \text{[TE-level coefficient variability]}\\
& && \text{[Participant-level coefficient variability]}\\
\begin{pmatrix} u_{k_0} \\ u_{1_i} \\ \vdots \\ u_{8_i} \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \Sigma_u \right)\\
\Sigma_u &= \begin{pmatrix} \rho_{u_0} & \rho_{u_0}\sigma_{u_{0_k}}\sigma_{u_1} & \cdots & \rho_{u_0}\sigma_{u_0}\sigma_{w_8} \\ \rho_{u_1}\sigma_{u_1}\sigma_{u_0} & \rho_{u_1} & \cdots & \rho_{u_1}\sigma_{u_1}\sigma_{u_8} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_8\sigma_{u_8}\sigma_{u_{0_k}} & \cdots & \cdots & \rho_{u_8} \end{pmatrix}\\
\sigma_{u_{0-8}} &\sim \mathcal{N}_+(1, 0.1)\\
\rho_u &\sim LKJcorr(2)\\
& && \text{[TE-level coefficient variability]}\\
\begin{pmatrix} w_{k_0} \\ w_{1_j} \\ \vdots \\ w_{3_j} \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \Sigma_w \right)\\
\Sigma_w &= \begin{pmatrix} \rho_{w_0} & \rho_{w_0}\sigma_{w_{0_k}}\sigma_{w_1} & \cdots & \rho_{w_0}\sigma_{w_0}\sigma_{w_3} \\ \rho_{w_1}\sigma_{w_1}\sigma_{w_0} & \rho_{w_1} & \cdots & \rho_{w_1}\sigma_{w_1}\sigma_{w_3} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_3\sigma_{w_3}\sigma_{w_{0_k}} & \cdots & \cdots & \rho_{w_3} \end{pmatrix}\\
\sigma_{w_{0-3}} &\sim \mathcal{N}_+(1, 0.1)\\
\rho_{w_{0-3}} &\sim LKJcorr(2)
\end{aligned} \tag{6}
$$

where: $\rho_{u_{0-8}}$ and $\rho_{w_{0-3}}$ indicate the correlation parameters between the by-participant and by-TE adjustments, respectively, $\sigma^2_{u_{0-8}}$ and $\sigma^2_{w_{0-3}}$ indicate the variance of the by-participant and by-TE variance of the adjustments, respectively, and $\mathcal{N}$ indicates a normal distribution, $\mathcal{N}_+$ indicates a truncated normal distribution with only positive values, and $LKJcorr$ indicates a LKJ correlation distribution (Lewandowski et al., 2009).

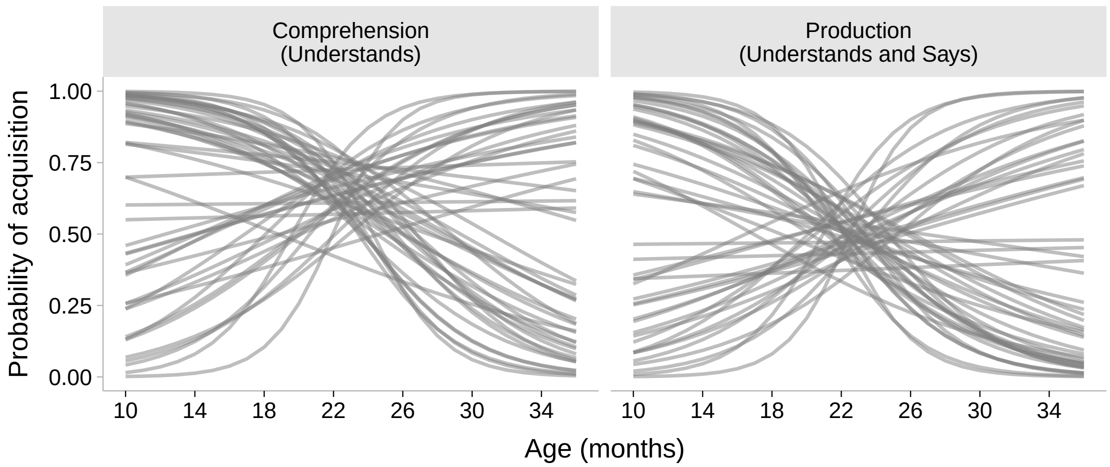

Figure 1: Expected prior-predicted mean.

Appendix C: convergence diagnostics

We used Stan (Carpenter et al., 2017) as the probabilistic language behind the estimation of our Bayesian models in this study, with `brms` as its R interface (Bürkner, 2017). This language implements the Markov Chain Monte-Carlo (MCMC) algorithm to explore the posterior distribution of the model. Broadly, this algorithm is used to iteratively sample the joint sampling space of the parameters to be estimated in the model, and compute, for each value sampled, its likelihood under some probability distribution previously defined. We run 4 MCMC chains, each 1000 iterations long each. The correct performance of this algorithm is critical to the quality of the statistical evidence to which the outcomes of the model lead.

One way to diagnose the behaviour of MCMC is to inspect whether the different MCMC chains (if more than one) have converged to a similar region of the posterior. The Gelman-Rubin diagnostic ($\hat{R}$ or R-hat Gelman & Rubin, 1992) provides a measure of chain convergence by comparing the variance within each chain *versus* the variance between each chain. Both are expected to be identical when chains have perfectly converged, so that $\hat{R} = 1$. Values lower than 1.01 are recommended, while values higher than 1.05 indicate that chains might have trouble converging and therefore the estimated parameters must be taken with caution. Figure 2 (A) shows the distribution of $\hat{R}$ values for the coefficients of the fixed effect of our models, which we used for statistical inference. Most values are lower than 1.01, and never higher than 1.05, which provides evidence of successful MCMC convergence.

Another diagnostic of good MCMC converge is the ratio of effective sample size to total sample size (N_{eff}/N), which indicates the proportion of samples in the chain that resulted from a non-divergent transition. Values closer to 1 are ideal, as they indicate that all posterior samples from the MCMC were used to estimate the posterior distribution of the parameter. Values larger than 0.1 are recommended. Figure 2 (B) shows the distribution of the effective sample sizes of the coefficients of the fixed effects in our models. Most values are larger than 0.1, although model 0 ($\mathcal{M}_0$) accumulates most effective sample sizes close to 0.1.

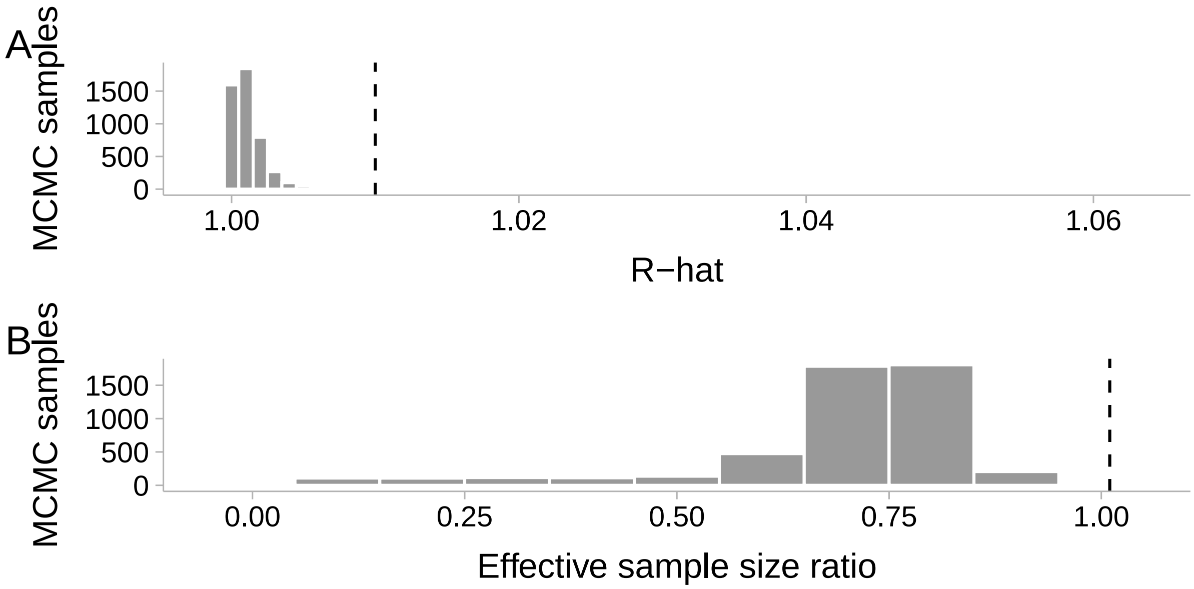

Figure 2: MCMC convergence diagnostic of all parameters in the model. A: distribution of the Gelman-Rubin (R-hat) scores. B: distribution of the ratio of effective sa

Posterior draws and bi-variate scatterplots

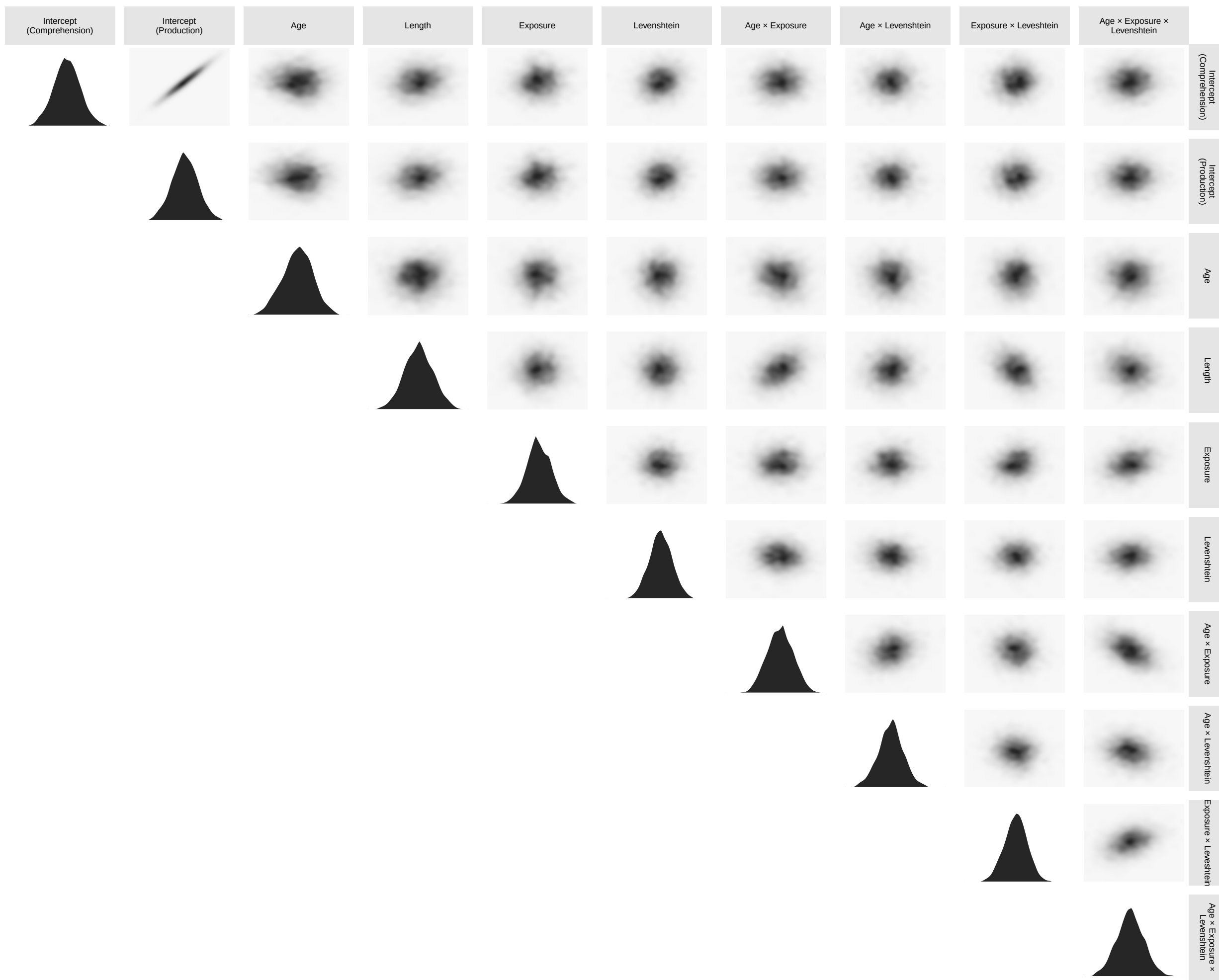

Figure 3: Marginal distribution and bi-variate scatterplot of posterior samples for the fixed regression coefficients in Model 3.

Posterior-predictive checks

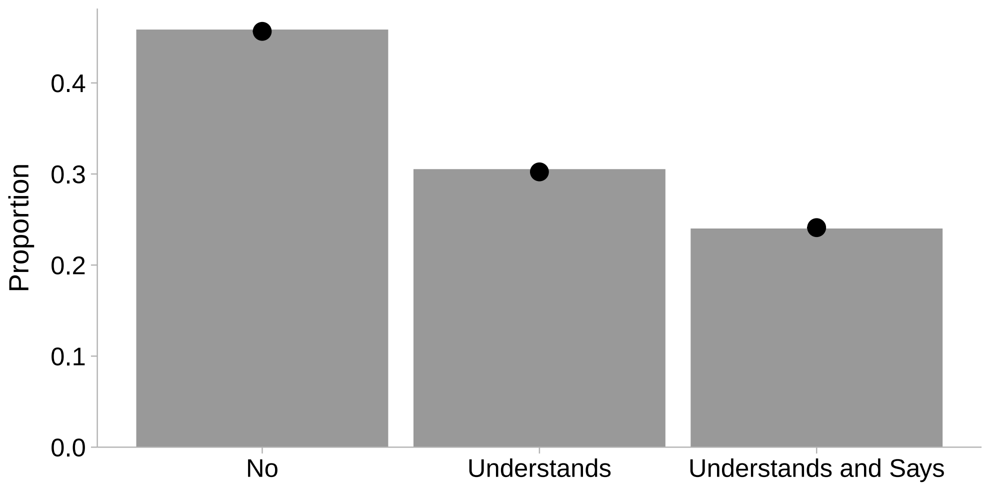

Figure 4: Model posterior predictive checks (PPC). Bars indicate the observed proportion of responses to each category (No, Understands, and Understands and Says). Blue dots and error bars represent the mean proportion of responses simulated from the posterior for each category, and its 95% interval.

Appendix D: frequency and language exposure as separate predictors

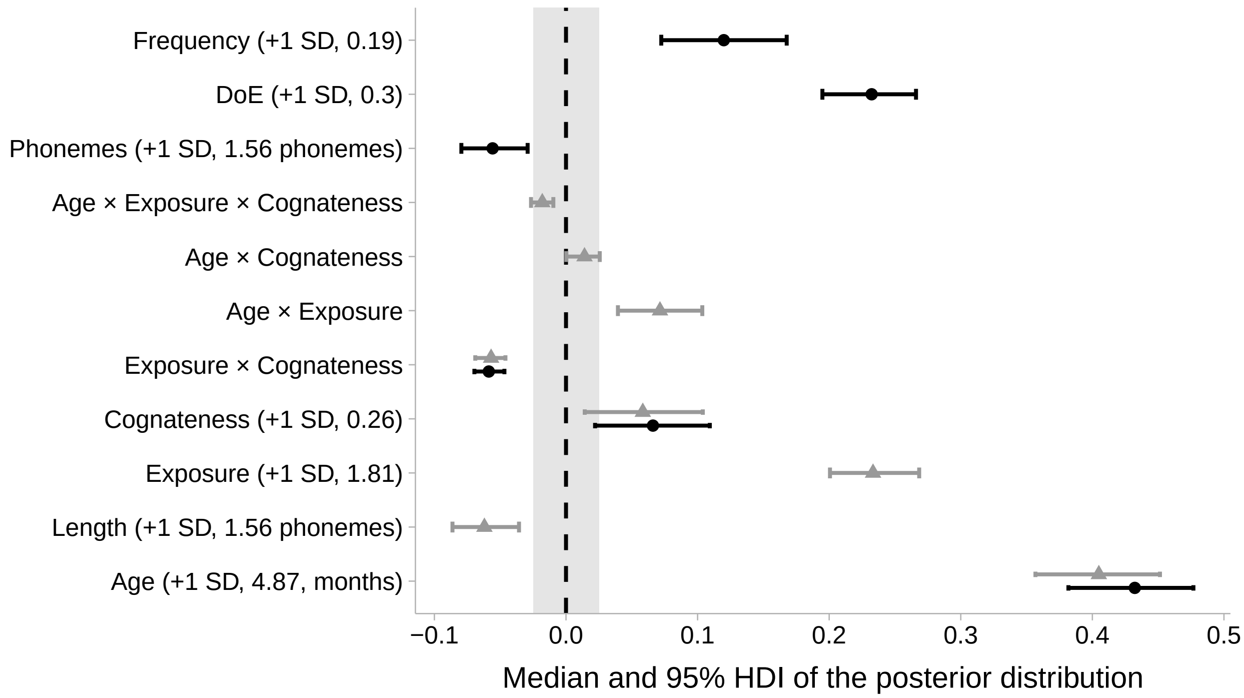

Figure 5: Posterior distribution of regression coefficients in the probability scale

Appendix E: frequency and language exposure as separate predictors

We define syllable frequency as the rate of appearance of individual syllables in the word-forms included in the Barcelona Vocabulary Questionnaire (BVQ). Each item corresponds to a Catalan or Spanish word, and has an associated phonological transcription in X-SAMPA format. These transcriptions are syllabified. Some examples:

Table 2: Sample of items included in the BVQ questionnaire and their syllabified SAMPA transcriptions in Catalan and Spanish

	Catalan			Spanish		
Translation	Item	X-SAMPA	Syllables	Item	X-SAMPA	Syllables
ball	futbol	fud"bO5	2	fútbol	"fut.bol	2
coffee	cafè	k@"fE	2	café	ka"fe	2
neighborhood	colònia	ko"lo.nja	3	colonia	ko"lo.nja	3
animal	Animals	"{n.I.m@lz	3	animales	a.ni"ma.les	4
family	família	fa"mi.lia	3	familia	fa"mi.lia	3
apple	poma	"po.m@	2	manzana	man"Ta.na	3
store	botiga	bu"ti.G@	3	tienda	"tjen.da	2
photo	càmera	"ka.m@.4@	3	cámara	"ka.ma.4a	3
box	caixa	"ka.S@	2	caja	"ka.xa	2
knife	ganivet	g@.ni"BEt	3	cuchillo	ku"tSi.Lo	3
tractor	tractor	t4@k"to	2	tractor	t4ak"to4	2
water (beverage)	aigua	"aj.Gw@	2	agua	"a.Gwa	2
giraffe	girafa	Zi"4a.f@	3	jirafa	xi"4a.fa	3
tray	safata	s@"Ba.t@	3	bandeja	ban"De.xa	3
ham	pernil	p@r"ni5	2	jamón (york)	"xa"mon	2

Most Catalan and Spanish words had two syllables, with Spanish words having three and four syllables more often than Catalan words. Less than 1% of the words included in the analyses presented in the main body of the manuscripts had five syllables. No words had more than five syllables (see Figure 6). We extracted lexical frequencies from the English corpora in the CHILDES database. Using the Catalan and Spanish corpora was not possible due to the low number of children and tokens included in the corpora.

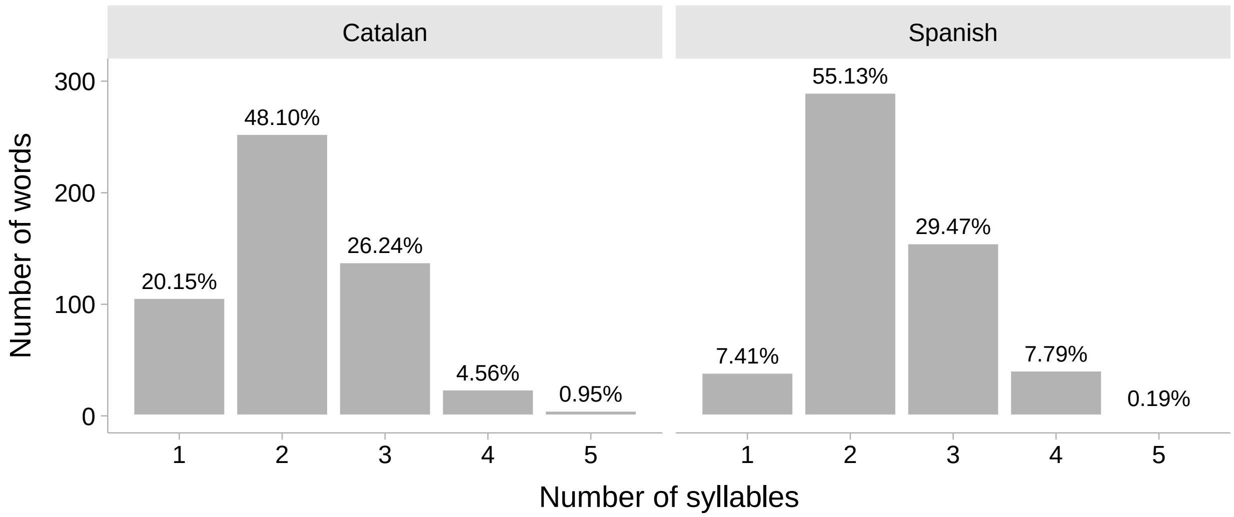

Figure 6: Distribution of the number of syllables in Catalan and Spanish

We now present how syllable frequencies were calculated. Every exposure to a word-form also counts as a exposure to each of the syllables that make up such word. Every time a child hears the word *casa* [house], they are exposed to the syllables *ca* and *sa*. Syllables that appear embedded in words with higher lexical frequency will also be more frequent. To compute the relative frequency of each syllable in Catalan and Spanish (i.e., how many times the syllables appears in every million words in Catalan or Spanish speech), we summed the relative lexical frequency in CHILDES of every word that contains such syllable in the corresponding language. Figure 7 shows the distribution of frequencies across syllables in Catalan and Spanish. In the log10 scale, syllable frequencies in Catalan and Spanish followed a slightly asymmetric distribution, with most syllables scoring around 1,000 counts per million, and a longer tail to the right of the distribution.

Figure 8 shows the top twenty most frequent syllables in Spanish. In both languages, most of the most frequent syllables followed a consonant-vowel (CV) structured. In Catalan, the most common nucleus in the most frequent syllables was /ə/ (most likely due to vowel reduction in unstressed syllables), followed by /u/. In Spanish, the most common nucleus in the most frequent syllables were /a/ and /o/.

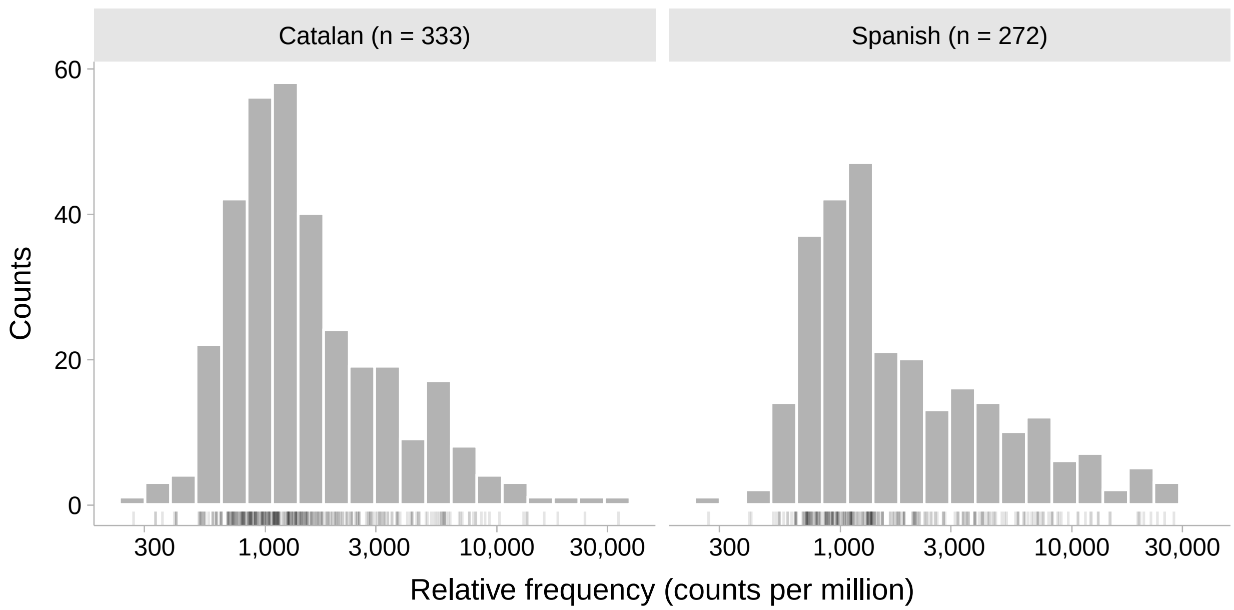

Figure 7: Distribution of apositional syllable frequencies in Spanish and Catalan

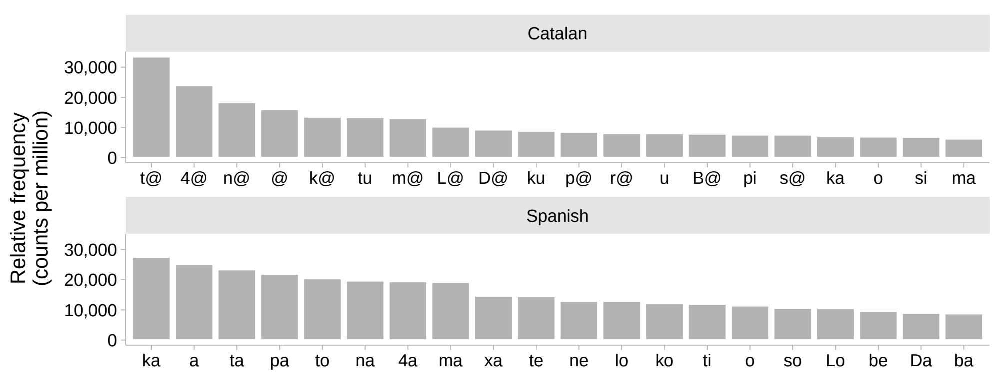

Figure 8: Top 20 most frequent syllables

To estimate the association between word-level syllabic frequency and cognateness while controlling for the number of syllables in the word, as words are expected to necessarily increase the syllabic frequency of the word), we fit a multilevel, Bayesian linear regression model with syllabic frequency (the sum of the syllabic frequency of the syllables in a word) as response variable, and the main effect of the number of syllables (*Syllables*) and *Cognateness* (Levenshtein similarity between a word and its translation equivalent) as predictors. We added translation equivalent-level random effects for the intercept and the main effect of *Syllables* (some translation pairs had a different number of syllables in each language). We used a Gaussian distribution to model syllabic frequency scores after standardising this variable and the predictors. We used a weakly informative prior for all parameters involved in the model (see Equation 7 for a formal equation of this model and its prior).

$$
\begin{aligned}
\text{Syllable frequency} &\sim \mathcal{N}(\mu, \sigma) \\
\mu &= (\beta_0 + u_{0_i}) + (\beta_1 + u_{1_i})\text{Syllables} + \beta_2\text{Cognateness} \\
\beta_{0-3} &\sim \mathcal{N}(0, 10) \\
u_{0-1_i} &\sim \mathcal{N}(0, \sigma_{u_i}) \\
\sigma_y &\sim \text{Exponential}(2) \\
\begin{pmatrix} u_{0_i} \\ u_{1_i} \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_u\right) \\
\Sigma_u &= \begin{pmatrix} \sigma_{u_0} & \rho_{u_0}\sigma_{u_0}\sigma_{u_1} \\ \rho_{u_1}\sigma_{u_1}\sigma_{u_0} & \sigma_{u_1} \end{pmatrix} \\
\sigma_{u_{0-1}} &\sim \mathcal{N}_+(1, 0.1) \\
\rho_u &\sim LKJcorr(2)
\end{aligned}
\tag{7}
$$

We fit this model running 4 sampling chains with 1,000 iterations each. Table 3 shows a summary of the posterior distribution of the fixed effects in the model. As expected, words with more syllables scored higher in syllabic frequency: all posterior draws for the regression coefficient of the main effect of this predictor fell outside the ROPE defined between -0.5 and +0.5 (β = 5.64, 95% HDI = [5.58, 5.71]). Keeping the number of syllables constant, the effect of cognateness was negligible: all of the posterior distributions of this predictor fell within the ROPE, providing evidence that the true value of the increment in syllabic frequency for every increase in cognateness is equivalent to zero ($\beta = 0.01$, 95% HDI = [-0.06, 0.07]).

Table 3: Posterior distribution of regression coefficients.

	Median	95% HDI	ROPE prob.
Intercept	16.09	[16.02, 16.16]	0.00%
Syllables (+1 SD, 0.802)	5.64	[5.58, 5.71]	0.00%
Cognateness (+1 SD, 0.24)	0.01	[−0.06, 0.07]	100.00%

Figure 9 shows the median posterior-predicted syllabic frequencies for words with one to four syllables, for the whole range of cognateness values. Overall, cognate words' syllabic frequency is equivalent to that of non-cognates. This suggests that the cognate facilitation effect in word acquisition reported in the present study is not the result from an association between cognateness and higher syllabic frequencies.

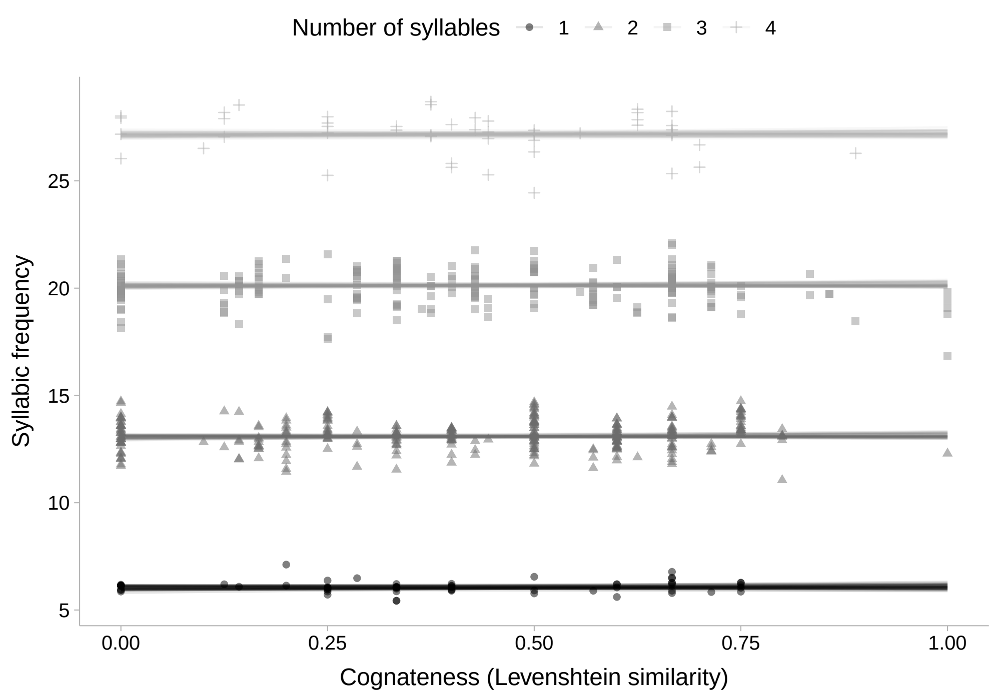

Figure 9: Posterior-predictions of the syllabic frequency model. Thicker lines indicate the median of the posterior predictions, and thinner lines indicate individual posterior predictions.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101.

Campbell, L., & Mixco, M. J. (2007). *A glossary of historical linguistics*. Edinburgh University Press.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan : A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

Floccia, C., Sambrook, T. D., Luche, C. D., Kwok, R., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot-Smith, K., Krott, A., Mills, D., Rowland, C., Gervain, J., & Plunkett, K. (2018). II: methods. *Monographs of the Society for Research in Child Development*, *83*(1), 30–42. https://doi.org/https://doi.org/10.1111/mono.12349

Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The growth of children's semantic and phonological networks: Insight from 10 languages. *Cognitive Science*, *44*(7), e12847. https://doi.org/10.1111/cogs.12847

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://www.jstor.org/stable/2246093

Heeringa, W. (2004). *Measuring dialect pronunciation differences using shtein distance* [PhD thesis]. University Library Groningen][Host].

Heeringa, W., & Gooskens, C. (2003). Norwegian dialects examined perceptually and acoustically. *Computers and the Humanities*, *37*(3), 293–315. https://doi.org/10.1023/A:1025087115665

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608. https://doi.org/10.1126/science.1736364

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics-Doklady*, *10*, 707–710.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001.

Loo, M. P. J. van der. (2014). The stringdist package for approximate string matching. *The R Journal*, *6*(1), 111–122. https://doi.org/10.32614/RJ-2014-011

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

McElreath, R. (2020). *Statistical rethinking: A bayesian course with examples in r and stan*. Chapman; Hall/CRC.

Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2019). Childes-db: A flexible and reproducible interface to the child language data exchange system. *Behavior Research Methods*, *51*(4), 1928–1941.

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in europe as based on levenshtein distance. *Bilingualism: Language and Cognition*, *15*(1), 157–166.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*(3), 281–284.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for british english. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.

Wells, J. C. (1995). *Computer-coding the IPA: A proposed extension of SAMPA*. *4*(28), 1995.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Addison-Wesley Press.